Balancing Cultural Specificity and Generalizability: Brief Qualitative Methods for Selecting, Adapting, and Developing Measures for Research With American Indian Communities

Emily E. Haroz[1], **Jerreed D. Ivanich**[2], **Allison Barlow**[1], **Victoria M. O'Keefe**[1], **Melissa Walls**[1], **Cindy Kaytoggy**[1], **Rose Suttle**[1], **Novalene Goklish**[1], **Mary Cwik**[1]

[1]Center for American Indian Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health

[2]Centers for American Indian and Alaska Native Health, Colorado School of Public Health

Abstract

Culturally appropriate, valid and reliable measures are critical to assessing how interventions impact health. There is a tension between measures for specific cultural settings versus more general measures that permit comparisons across samples. We illustrate a feasible approach to measurement selection, adaptation and testing for a study of brief interventions to prevent suicide among American Indian youth ages 10–24. We used a modified Nominal Group Technique (NGT) with $N = 7$ Apache Community Mental Health Specialists (CMHS') to elicit priority impacts of interventions under study. We then tested the reliability and validity in $N = 93$ youth at baseline. The NGT results included selection of alternative measures, item removal and addition, and creation of a local well-being index. Measurement testing indicated excellent to good internal consistency (α: 0.82–0.96) and strong construct validity. Study results demonstrate a feasible approach to balancing cultural specificity and generalizability while producing valid and reliable measures to use in an intervention trial.

Keywords

Native American; mixed methods; validity; reliability; measures

Author Manuscript

All scientific study is built on an assumption of valid and reliable measurement. In prevention and intervention research, measuring the impact of our interventions with measures only assumed to be valid could bias our results (Gottfredson et al., 2015). This bias could yield over confidence in the results of interventions or reduce our ability to see intervention effects (Wilson & Lipsey, 2001). Large investment in intervention development and testing without commensurate investment in validating measures across diverse populations is limiting prevention science's capacity to promote health equity. This challenge is the greatest in communities that have been historically overlooked in the validation of established measures, such as American Indian or Alaska Native (AIAN) populations. These communities typically have fewer resources, are underserved, and are in highest need of effective prevention interventions.

The purpose of this paper is to describe our process to select, adapt, and develop measures for use in a large study focused on testing the impact of brief interventions aimed at reducing suicide risk and promoting resilience among Native American youth ages 10–24 (O'Keefe et al., 2019). Our approach reflects understanding of a tension that exists in all research across populations—that is, whether the phenomena being measured is universal to all peoples, or the meaning and expression of constructs under consideration are specific to each community and cultural setting. Whether the phenomenon pertains to physical, mental, emotional, or social states of health matters. For example, measuring whether a person has been infected with a virus does not require the redevelopment of existing assessment tools for each specific cultural context as it is detecting a biological process that is universal to human biology. However, measuring a person's level of depression, social support or resilience demands particular attention to contextual and cultural meaning and expression as we are often measuring these latent constructs with items that may or may not represent the phenomenon the same in each setting (Burkey et al., 2018; Doty et al., 2018; Haroz, Bass, et al., 2014; Haroz, Ritchey, et al., 2017). For example, Jayawickreme et al. (2012), in their work in Sri Lanka, tested the hypothesis that psychometric instruments that were culturally adapted predicted functional impairment in a non-Western population better than translated established instruments. They found that the culturally adapted instruments predicted functional impairment above and beyond what was predicted by the nonadapted instruments. Haroz, Bass, et al. (2017), found similar findings: a scale that included items based on qualitative studies from around the world that are not included in standard depression measures (International Depression Symptom Scale, IDSS), equally predicted a diagnosis of major depression as an un-adapted Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) in Myanmar. However, the IDSS also predicted functional impairment among individuals with depression even after controlling for the the variance explained by the PHQ-9.

Prevention intervention science that is aimed at overcoming behavioral and mental health disparities across diverse settings requires balancing the need to practically incorporate measurement adaptations or key indicators of priority to the local context, while ensuring fidelity to the measure's demonstrated psychometrics to enhance generalizability. The urgency of particular health issues such as youth suicide also adds pressure on both sides of this balance in terms of: (a) the importance of unlocking the unique risk and protective

factors in settings disproportionately affected by suicide, and (b) finding measures and interventions that can scale rapidly to others communities in high need.

This tension between measurement approaches that allow for cross-cultural comparisons versus approaches that are specific to each community and cultural setting, is present in all research, but has been a large focus of research with AIAN communities. Much of the research focusing on AIAN has taken more of an etic approach—assuming the universality of the underlying phenomenon being measured. Examples of this are numerous and often involve careful selection and use of measurement scales developed in non-Native populations (Mullany et al., 2012; Walls et al., 2007; Whitbeck et al., 2009). Less developed are emic approaches—developing measures specific to one community or group because the underlying phenomenon is thought to only be understood within the specific cultural context. A widely used example of an emic approach to measurement in Native American research is Whitbeck et al.'s work on conceptualizing and measuring Historical Trauma (Whitbeck et al., 2004). Other scholars working with Native communities are increasingly developing new emic measures (Allen et al., 2019; Fok et al., 2012; Mohatt et al., 2011; Oetting & Beauvais, 1990–1991). These approaches parallel efforts more broadly in the global mental health field. With recognition that translation and back-translation of measurement instruments is not sufficient to ensure measure validity, many researchers have turned to brief ethnographic methods that can be used to inform measure selection and adaptation (Bolton & Tang, 2002; Haroz, Bass, et al., 2014; Weaver & Kaiser, 2015). For example, Doty et al. (2018), used data from free listing and key informant interviews to adapt several existing measures for mental health outcomes and created unique indexes of impaired functioning specific to internally displaced and veteran's populations in Ukraine.

This paper presents an example of processes that leverages both etic and emic approaches with an American Indian community. We present the methods and outcomes of our process to select, adapt, and supplement outcome measures related to suicidal ideation and resilience —the key targets of our trial, followed by preliminary psychometric analysis of the newly adapted or designed measures with $N = 93$ youth enrolled in the trial at baseline. The process we undertook is both feasible and replicable, with the potential to serve as a brief method that can practically incorporate outcomes of priority to the local context and community, while balancing fidelity to other measures that enhance generalizability. While we focus specifically on our work with AIAN communities, these methods and approaches are widely applicable to many communities both within the United States and more broadly. Many standard measures were not developed with a diverse range of racial and ethnic populations. The methods provided in this paper may help enhance cultural fit of our measurement tools across diverse populations.

Method

Instrument Development and Adaptation

University-based and tribal community researchers collaborated to select, adapt, and test outcome measures for this intervention trial. This study was approved by the Johns Hopkins Institutional Review Board (#8138) and by the Health Board of the White Mountain Apache Tribe. This study was not preregistered. We were guided by a grounded theory approach

to adapt and develop the study instruments. Grounded theory involves qualitative methods as a means to discover new information and generate hypotheses (Charmaz & Belgrave, 2015). Broad open-ended questions are used to elicit participants' thoughts and beliefs without being influenced by a priori hypotheses or preestablished theories. Specifically, we convened a focus group discussion (FGD) of stakeholders and used a modified Nominal Group Technique (NGT) to elicit and prioritize key stakeholders' thoughts and beliefs about the changes that resulted in youth from their participation in suicide prevention efforts (process described below).

The focus group was held as part of a training on the research procedures for the larger clinical trial which is testing two brief interventions for youth ages 10–24 with recent suicide attempts, suicide ideation, or binge drinking and suicide ideation in partnership with the White Mountain Apache Tribe (O'Keefe et al., 2019). The first psychoeducational-based intervention is provided by Apache case managers, focused on reducing suicide risk, and was previously adapted and tested in the same community (i.e., New Hope; Cwik et al., 2016). The second culturally based intervention is delivered by community Elders with support from Apache Community Mental Health Specialists (CMHS'), and was developed and implemented in the same community (Cwik et al., 2019). All trial participants receive case management from Apache CMHS'. Primary outcomes are reductions in suicide ideation and increases in resiliency.

The key stakeholders for the FGD were $N = 7$ Apache CMHS' who had previous experience delivering both interventions. These CMHS' were from the community, had worked on local suicide prevention efforts for several years, and were familiar with local beliefs, norms, and culture, as well as with implementing research protocols. In the FGD, the main outcome measures that had been proposed for the larger clinical trial: the Suicide Ideation Questionnaire (SIQ; Reynolds, 1988) and the Prince-Embry Resilience Scales (Prince-Embury, 2008), were reviewed. The Suicide Ideation Questionnaire (SIQ) has different versions based on an individual's age (SIQ and SIQ-Junior). The trial sample consists of individuals that are eligible for both versions of the SIQ, thus we used the fifteen questions that the SIQ and SIQ-Junior have in common (Reynolds & Mazza, 1999). Possible response options include, 0 "I never had this thought," 1 "I had this thought before but not in the past month," 2 "about once a month," 3 "couple times a month," 4 "about once a week," 5 "a couple of times a week," and 6 "almost every day" for a possible score range of 0–90. We used three subscales of the Prince-Embry Resilience Scales—sense of mastery (composed of 20 items), sense of relatedness (24 items), and emotional reactivity (20 items; Prince-Embury, 2008). The sense of mastery scale assesses whether an individual feels a sense of optimism about life, as well as one's adaptability and problem-solving abilities. Sense of relatedness evaluates one's ability to interact with others and/or tolerate others. The emotional reactivity scale assesses one's sensitivity to negative events. Each item has four response choices ranging from 0 "Not at all" to 5 "Almost Always." All items on the emotional reactivity scale are reverse coded to provide a score of how emotionally proactive youth are.

We also reviewed all items on secondary measures including: (a) Demographics; (b) The Centers for Epidemiologic Studies of Depression Revised 10-item version (Haroz, Ybarra,

& Eaton, 2014); (c) Apache Hopefulness Scale (Hammond et al., 2009); (d) Youth Risk Behavior Survey Substance Use Items (Kann et al., 1993); (e) UPPS Impulsive Behavior Scale (Whiteside & Lynam, 2001); (f) Hemingway Measure of Adolescent Connectedness (Karcher, 2008); (g) Voices of Indian Teens Cultural Issues and Interest (Moran et al., 1999); and (h) the Rosenberg Self Esteem Scale (Rosenberg, 2015).

During the FGD items were identified on the SIQ and the Prince-Embry Resilience Scales as problematic if they were not relevant in the local context or if the wording needed to be changed to ensure the item could be meaningfully interpreted in the Apache community. If the item was not relevant, it was removed. If the wording needed to be changed, we revised the language until we arrived at a consensus on which wording best captured the concept while also being understood locally.

During the FGD we utilized NGT to guide the discussion and posed two questions to the FGD participants: (a) What are all the changes that a youth experiences because they participated in New Hope? and (b) What are all the changes that a youth experiences because they participated in the Elders Resilience Curriculum? First, we asked the CMHS' to independently make a list of all the changes resulting from each intervention they could think of on their own. Second, we asked CMHS' to go back again individually and select the top five changes resulting from each intervention that they felt were most important. Five was selected as a feasible number that would balance prioritization complexity and the overall length of the final list (e.g., maximum number would be 35 unique changes for each intervention based on having seven FGD participants who could prioritize five changes). Third, we asked each person to share their top five changes with the larger group to create a comprehensive list of the most important changes they felt youth experienced as a result of each intervention. Items that reflected the same change/concept were grouped together based on group consensus. If consensus could not be reached, items were kept separate.

After completion of the FGD, we compared the changes identified as resulting from each intervention to the items and scales in the study measures. Where the NGT results overlapped with an item on the draft instrument battery, we documented it. Where the NGT results did not reflect our proposed items or instruments, we added to or replaced the existing items with new questions or scales that would better capture the relevant local worldviews. Finally, we created a local index of well-being with the remaining changes that were identified by the FGD, but still not captured in any existing instruments. We limited the questions on this index to items believed to be amenable to intervention. The resulting instrument battery therefore incorporated feedback on wording changes, local item relevance, consistency with locally observed changes, and a new index aimed at capturing changes in youth that were locally meaningful.

Instrument Testing

To examine whether these adaptations and new additions performed adequately, we examined their performance in the first $N = 93$ participants enrolled in the trial (O'Keefe et al., 2019). Performance was measured by examining Cronbach's α (reliability; Bland & Altman, 1997), distribution of baseline score (variability), and construct validity using a hetero-trait monomethod correlation matrix (i.e., convergent and discriminant validity;

Campbell & Fiske, 1959). Cronbach alphas greater than 0.90 are considered to have excellent internal consistency; 0.80–0.90 good; and 0.70–0.80 acceptable; while below 0.70 are considered unacceptable (Tavakol & Dennick, 2011). For construct validity, we examined the relationship of the adapted versions of the SIQ, Resiliency Scales, and the wellness index to other scales used in the study including the Children's Hope Scale (CHS; Snyder et al., 1997) and the Centers for Epidemiologic Study of Depression—Revised 10 item scale (CESD-R-10; Haroz, Ybarra, & Eaton, 2014). Using a hetero-trait monomethod approach to construct validity we would expect: (a) our adapted version of the SIQ to be significantly negatively correlated with the Resilience scales, the CHS, and the well-being index (i.e., discriminant validity) to be positively correlated with the CESD-R-10 (i.e., convergent validity); and (b) our adapted versions of the Resiliency scales and newly created well-being index to be significantly negatively correlated with the SIQ and the CESD-R-10 and positively correlated with the CHS and each other.

This study was not preregistered. Data from this project are available upon reasonable request and approval from the Tribal partner.

Results

Results From the Item Review of Each Measure

Suicide Ideation—Three items were removed from the Suicide Ideation Questionnaires (SIQs; Reynolds & Mazza): "I thought about people dying," "I thought about writing a will," and "I thought about telling people I plan to kill myself." These items were removed based on recommendations from the focus group discussion. Apache CMHWs stated that talking about death in the way the SIQ items were worded was not culturally appropriate nor common. In addition, FGD participants shared that drafting wills is not a common practice on the reservation. In a prior study assessing validity, reliability, and factor structure of the SIQ with WMAT youth, the same three items were found to be the worst performing items and removed to improve model fit in a Confirmatory Factor Analysis (Hill et al., 2018).

Resilience—Based upon FGD results with Apache CMHWs, several items required revised language to ensure interpretability of local meaning and some items were removed due to not being relevant on the Prince-Embry Resilience Scales (Prince-Embury, 2008). For the Sense of Mastery subscale, all items were retained, and two items required rewording. For the Sense of Relatedness subscale, three items were removed, as Apache CMHWs in the FGD recommended they were not sensitive enough to change based on the interventions being studied (i.e., "I can meet new people easily," "I can make friends easily," and "I have a good friend."). Four items on the Sense of Relatedness subscale needed rewording to be relevant to the local context. In addition, Apache CMHWs believed that the Sense of Relatedness subscale was incomplete for the local setting and added six items: (a) "When I am upset I know what to do to make me feel better"; (b) "When I am upset I know who to talk to who can help me"; (c) "When I am upset I can control my behavior"; (d) "When I am upset I know it will be better tomorrow"; (e) "When I am upset I know it won't last forever"; and (f) "When I am upset I know prayer can help me." For the Emotional Reactivity subscale all items were retained; however, 15 items needed rewording

for the local context. Finally, Apache CMHWs recommended that scoring categories on the Prince-Embry Resilience Scales be reduced from five to four categories and that qualitative descriptors of each category reflect modifications to item wording (e.g., "Never" was changed to "Not at all").

Results of the Modified Nominal Group Technique

FGD participants noted several changes they have observed among youth who have received the New Hope intervention, including: (a) gaining skills at managing anger and impulsivity; (b) increasing self-esteem; (c) becoming more hopeful; and (d) understanding their network of social support. During the FGD, Apache CMHWs also reported changes they have observed among youth who have received the Elder's Resilience Curriculum intervention, including: (a) increased sense of cultural identity and belonging; (b) gaining confidence; and (c) being more respectful toward others. See Table 1 for the final list of observed changes youth experience because they participated in the New Hope and/or Elders' Resilience Curriculum interventions. Following the FGD, we compared changes identified during the NGT process to scales on the draft assessment battery. For observed changes in youth identified during NGT that were not reflected in the instrument battery, we added stars to signify a need to find other instruments or create new items/instruments to capture these changes. All changes to the assessment battery and rationale for these changes are outlined in Table 2.

NGT results revealed that there were observed changes among youth who received the New Hope or Elders' Resilience Curriculum interventions that did not reflect items or instruments in the draft instrument battery. Therefore, we worked collaboratively with Apache CMHWs to add 11 items as an Index of Local Indicators of Well-Being (Table 3). Response options for the Index of Local Indicators of Well-Being was created to be generally consistent with other study scales (0 = *not at all*; 3 = *a lot*). We refrain from calling this a scale, as it is unclear whether the items tap into the same underlying latent construct or represent items in different latent traits that are combined into the same list. Despite this limitation, these indicators represent positive changes that our community partners identified as important impacts of the New Hope and Elders' Resilience Curriculum interventions and therefore critical to monitor throughout the Sequential Multiple Assignment Randomization Trial (SMART) and future research evaluating these interventions. The process of generating Local Indicators of Well-Being represents an etic approach to measure development, as it derives from open-ended inquiry to inform locally important outcomes that are distinct from existing measures.

Testing Psychometrics of Indicators

Participant Characteristics—The majority of participants were female (70%) and slightly less than half (42%) of the sample were between the ages of 10–14. All participants experienced a recent suicide attempt, suicide ideation, or binge substance use with suicide ideation to be eligible to participate in the larger study (O'Keefe et al., 2019). Most participants in the present study sample experienced recent suicide ideation (62%), while one-third of participants had a recent suicide attempt (30%), and 8% of participants

experienced recent binge substance use with suicide ideation. See Table 4 for sample demographics.

Results of Psychometric Analyses—Table 5 provides the average scores, standard deviations, score ranges, and Cronbach's α for suicide ideation (SIQ), resilience (Prince-Embry Resilience Scales), and local well-being (Index of Local Indicators of Well-Being) measures. The adapted SIQ showed excellent internal consistency (α = 0.96). All subscales of the adapted Resilience scales showed excellent or good internal consistency: Sense of Mastery subscale (α = 0.91), Sense of Relatedness subscale (α = 0.88), and Emotional Reactivity subscale (α = 0.91). The average score on the Index of Local Indicators of Well-Being for youth in our sample was 23.74 and internal consistency was good (α = 0.82). The slightly lower α for the well-being index is to be expected, as it is an index rather than a scale. Therefore, we do not expect these items to be unidimensional.

Table 6 provides the correlation matrix of our measures used to examine construct validity. As expected, our adapted version of the SIQ was significantly negatively correlated with all of the Prince-Embry Resilience subscales (Mastery: $p < .001$; Relatedness: $p < .01$; and Emotional Reactivity: $p < .001$) and with hope ($p < .001$), and significantly positively correlated with depression scores ($p < .001$). The adapted versions of the resilience scales also performed as expected and were significantly negatively correlated with depression scores (Mastery: $p < .001$; Relatedness: $p < .001$; and Emotional Reactivity: $p < .001$), and significantly positively correlated with hope (Mastery: $p < .001$; Relatedness: $p < .001$; and Emotional Reactivity: $p < .01$) and the local index of well-being (Mastery: $p < .001$; Relatedness: $p < .001$; and Emotional Reactivity: $p < .01$). Finally, our local index of well-being was not significantly correlated with the SIQ, but significantly negatively correlated with depression ($p < .05$), and significantly positively correlated with hope ($p < .001$) and the resilience scales (Mastery: $p < .001$; Relatedness: $p < .001$; and Emotional Reactivity: $p < .01$). Taken together, this suggests that our adapted versions of the SIQ and the resilience scales, and our qualitatively derived index of local indicators of well-being, demonstrate both discriminant and convergent construct validity.

Discussion

Measurement is fundamental to all intervention research. Often measures are developed in one population and then applied to another. This practice has the potential to add error and weaken our conclusions about if and how interventions may work. In this study, we aimed to use brief qualitative approaches to select, adapt, and/or develop measures for use in a trial testing the impact of two brief interventions for American Indian youth at risk for suicide. Our results demonstrate a feasible process to identify locally relevant intervention effects and leverage this information to select and adapt measures developed in non-Native populations for use with our target sample. Results from the psychometric analysis show that measures adapted and created through this process had a high level of internal consistency and demonstrate construct validity.

The efforts detailed in this manuscript add a level of structure and transparency to developing measures with community partners often not found in the literature. We intended

to lay out our methods in a step-by-step fashion so that others in the field could replicate or expand upon this process in their own work. Our approach focused on identifying items on scales that operated differently in our context using qualitative inquiry to inform item inclusion and revision. It is important to note that there are other approaches that may be relevant to item-level performance cross-culturally, including latent content analysis (Kleinheksel et al., 2020) and work that has been done and the use of Differential Item Function (DIF) analysis (Haroz et al., 2016). Measurement selection, adaptation, and development mostly remains a black box in research with Native communities. Many other authors have produced useful scales that work well in these communities, but the details for how these were developed are sometimes underspecified. This article builds on the work of these scholars by detailing the process involved in measurement development—a needed step moving forward with tribal partners with implications for replication.

Recently, Walls et al. (2019) proposed a framework for thinking strategically about the tension between tailored and common measurement approaches throughout the selection, adaptation, creation, and implementation of measures in research. The framework emphasizes a cycle of measurement development completed in partnership with AIAN community members and researchers (Walls et al., 2019). In the center of the framework is a rotating wheel reflecting the degree of cultural specificity that is required across the measurement development cycle going from conceptualization, to operationalization, to implementation and finally to interpretation. Community-researcher partnerships then determine the alignment of the measurement cycle and the specificity wheel that meets the local needs and aims of research project. Our work in this paper demonstrates how community-researcher partnerships "drove" our decisions about measurement. We focus here mainly on the conceptualization and operationalization of measures. Notably, some measures from the existing literature were deemed appropriate for this context, while other measures had to be adapted, replaced altogether, or created from scratch due to a dearth of measures that captured locally relevant concepts. This finding underscores the value of a range of measurement and cultural specificity, indicating that no one single solution will be appropriate for all studies.

Several notable results about the content of the assessments emerged from the brief qualitative measurement adaptation process, which have important implications for other groups evaluating similar constructs with Indigenous populations. Regarding suicide ideation, our research confirmed what has been found previously in a quantitative study— that items related to thinking about people dying, writing a will and thinking about telling others about your suicide ideation do not seem to be relevant in this sample (Hill et al., 2018). Suicide ideation is an often-measured construct in Indigenous populations because of the significance of this public health problem, as well as national and international efforts toward a "Zero Suicide" model of screening and risk assessment in medical settings (Labouliere et al., 2018). Administrators, clinicians, and researchers should proceed with caution before implementing evidence-based suicide assessments unless the measures have been validated with Indigenous populations or they adapt them for their own communities.

For resiliency, some of the subscales did not need many adaptions based on our process (e.g., sense of mastery), whereas others did (e.g., emotional reactivity and sense of

relatedness). These findings may be because as researchers we are just beginning to understand and refine how we measure resiliency, as well as the importance of culture to worldviews about social and emotional domains (Haroz, Bass, et al., 2017; Kirmayer et al., 2017; Liu et al., 2017; Ungar & Liebenberg, 2011). Finally, our study identified local indicators of well-being which included important themes that did not appear to be covered adequately by the other measures of mental health included in the study assessment battery, what might be referred to as daily functioning and connectedness, in particular to Elders.

Limitations

The results of this study should be interpreted with the following limitations in mind. First, our NGT process involved a small sample of CMHS' who while from the community, did not necessarily represent the full spectrum of opinions that could have been included. Other research has used similar processes (e.g., Free Listing; Key Informant Interviews) with other stakeholder groups, such as the target population for an intervention under study or those who have already received the intervention (Haroz, Bass, et al., 2014; Kaiser et al., 2013; Mazzuca et al., 2019). Second, our psychometric testing is limited because of the cross-sectional nature of the data and being embedded in baseline data collection for a larger trial. Reliability and validity testing of selected, adapted, or created measures prior to use in a research study is recommended over our strategy, yet is not always feasible given existing resources and funding timelines for studies, particularly for populations that are hard to recruit such as youth at risk for suicide. A carefully defined measurement validation could examine other, perhaps more rigorous forms of reliability such as test–retest reliability, criterion validity, and incremental validity. Finally, we did not evaluate cross-cultural measurement equivalence as it was beyond the scope of this study. To fully compare scores on the measures presented in this manuscript to the scores from a different population utilizing these same measures would require testing of configural, metric, and scalar invariance across population groups. Finally, it is important to note that we do not have data on whether our process changed measure interpretation or performance compared to the standard measures. However, data from other research would suggest that culturally adapted or created scales may perform better than standard measures used in different cultures and contexts from which they were developed (Greenfield et al., 2015; Haroz, Bass, et al., 2017; Jayawickreme et al., 2012).

Conclusion

In this paper we demonstrated a feasible approach to selecting, adapting, and generating measures for use in an evaluation of two brief interventions for American Indian youth at risk of suicide. Working through a strong community-researcher partnership, we aimed to balance cultural relevance and specificity with principals of generalizability to form an assessment battery that includes scales and items adapted or created for the local context. Moreover, these scales when used with youth performed as expected—demonstrating internal consistency, ability to capture variation in the sample, and construct validity. Ultimately, the methods described here are highly replicable and produced valid and reliable measures for use in a research trial with our target population. These methods are also feasible and widely applicable, including for work with any cultural or ethnic group in which standard measures may not fully fit the cultural context.

Acknowledgments

References

Allen J, Rasmus SM, Fok CCT, Charles B, Trimble J, Lee K, & Team Q (2019). Strengths-based assessment for suicide prevention: Reasons for life as a protective factor from Yup'ik Alaska Native youth suicide. Assessment, 28(3), 709–723, 10.1177/1073191119875789. [PubMed: 31538813]

Bland JM, & Altman DG (1997). Cronbach's alpha. BMJ (Clinical Research Ed.), 314(7080), Article 572. 10.1136/bmj.314.7080.572

Bolton P, & Tang AM (2002). An alternative approach to cross-cultural function assessment. Social Psychiatry and Psychiatric Epidemiology, 37(11), 537–543. 10.1007/s00127-002-0580-5 [PubMed: 12395144]

Burkey MD, Adhikari RP, Ghimire L, Kohrt BA, Wissow LS, Luitel NP, Haroz EE, & Jordans MJ (2018). Validation of a cross-cultural instrument for child behavior problems: The disruptive behavior international scale–Nepal version. BMC Psychology, 6(1), Article 51. 10.1186/s40359-018-0262-z [PubMed: 30390713]

Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81–105. 10.1037/h0046016 [PubMed: 13634291]

Charmaz K, & Belgrave LL (2015). Grounded theory. In The Blackwell encyclopedia of sociology. Wiley. 10.1002/9781405165518.wbeosg070.pub2

Cwik M, Goklish N, Masten K, Lee A, Suttle R, Alchesay M, O'Keefe V, & Barlow A (2019). "Let our apache heritage and culture live on forever and teach the young ones": Development of the elders' resilience curriculum, an upstream suicide prevention approach for American Indian youth. American Journal of Community Psychology, 64(1–2), 137–145. 10.1002/ajcp.12351 [PubMed: 31313327]

Cwik MF, Tingey L, Lee A, Suttle R, Lake K, Walkup JT, & Barlow A (2016). Development and piloting of a brief intervention for suicidal American Indian adolescents. American Indian and Alaska Native Mental Health Research, 23(1), 105–124. 10.5820/aian.2301.2016.105 [PubMed: 28562844]

Doty SB, Haroz EE, Singh NS, Bogdanov S, Bass JK, Murray LK, Callaway KL, & Bolton PA (2018). Adaptation and testing of an assessment for mental health and alcohol use problems among conflict-affected adults in Ukraine. Conflict and Health, 12(1), Article 34. 10.1186/s13031-018-0169-6 [PubMed: 30127843]

Fok CCT, Allen J, Henry D, Mohatt GV, & the People Awakening Team. (2012). Multicultural mastery scale for youth: Multidimensional assessment of culturally mediated coping strategies. Psychological Assessment, 24(2), 313–327. 10.1037/a0025505 [PubMed: 21928912]

Gottfredson DC, Cook TD, Gardner FE, Gorman-Smith D, Howe GW, Sandler IN, & Zafft KM (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. Prevention Science, 16(7), 893–926. 10.1007/s11121-015-0555-x [PubMed: 25846268]

Greenfield BL, Hallgren KA, Venner KL, Hagler KJ, Simmons JD, Sheche JN, Homer E, & Lupee D (2015). Cultural adaptation, psychometric properties, and outcomes of the native American spirituality scale. Psychological Services, 12(2), 123–133. 10.1037/ser0000019 [PubMed: 25961648]

Hammond VL, Watson PJ, O'Leary BJ, & Cothran DL (2009). Preliminary assessment of Apache hopefulness: Relationships with hopelessness and with collective as well as personal self-esteem. American Indian and Alaska Native Mental Health Research: The Journal of the National Center, 16(3), 42–51. 10.5820/aian.1603.2009.42 [PubMed: 20052632]

Haroz EE, Bass J, Lee C, Oo SS, Lin K, Kohrt B, Michalopolous L, Nguyen AJ, & Bolton P (2017). Development and cross-cultural testing of the International Depression Symptom Scale (IDSS): A

measurement instrument designed to represent global presentations of depression. Global Mental Health (Cambridge, England), 4, Article e17. 10.1017/gmh.2017.16

Haroz EE, Bass JK, Lee C, Murray LK, Robinson C, & Bolton P (2014). Adaptation and testing of psychosocial assessment instruments for cross-cultural use: An example from the Thailand Burma border. BMC Psychology, 2(1), Article 31. 10.1186/s40359-014-0031-6 [PubMed: 25685351]

Haroz EE, Bolton P, Gross A, Chan KS, Michalopoulos L, & Bass J (2016). Depression symptoms across cultures: An IRT analysis of standard depression symptoms using data from eight countries. Social Psychiatry and Psychiatric Epidemiology, 51(7), 981–991. 10.1007/s00127-016-1218-3 [PubMed: 27083900]

Haroz EE, Ritchey M, Bass JK, Kohrt BA, Augustinavicius J, Michalopoulos L, Burkey MD, & Bolton P (2017). How is depression experienced around the world? A systematic review of qualitative literature. Social Science & Medicine, 183, 151–162. 10.1016/j.socscimed.2016.12.030 [PubMed: 28069271]

Haroz EE, Ybarra ML, & Eaton WW (2014). Psychometric evaluation of a self-report scale to measure adolescent depression: The CESDR-10 in two national adolescent samples in the United States. Journal of Affective Disorders, 158, 154–160. 10.1016/j.jad.2014.02.009 [PubMed: 24655780]

Hill K, Van Eck K, Goklish N, Larzelere-Hinton F, & Cwik M (2018). Factor structure and validity of the SIQ-JR in a Southwest American Indian tribe. Psychological Services, 17(2), 207–216. 10.1037/ser0000298 [PubMed: 30475042]

Jayawickreme N, Jayawickreme E, Atanasov P, Goonasekera MA, & Foa EB (2012). Are culturally specific measures of trauma-related anxiety and depression needed? The case of Sri Lanka. Psychological Assessment, 24(4), 791–800. 10.1037/a0027564 [PubMed: 22429206]

Kaiser BN, Kohrt BA, Keys HM, Khoury NM, & Brewster ART (2013). Strategies for assessing mental health in Haiti: Local instrument development and transcultural translation. Transcultural Psychiatry, 50(4), 532–558. 10.1177/1363461513502697 [PubMed: 24067540]

Kann L, Warren W, Collins JL, Ross J, Collins B, & Kolbe LJ (1993). Results from the national school-based 1991 Youth risk behavior survey and progress toward achieving related health objectives for the nation. Public Health Reports, 108(1), 47–67. [PubMed: 8210275]

Karcher MJ, Holcomb M, & Zambrano E (2008). Measuring adolescent connectedness: A guide for school-based assessment and program evaluation. In Coleman HLK & Yeh C (Eds.), Handbook of school counseling (pp. 649–669). Lawrence Erlbaum.

Kirmayer LJ, Gomez-Carrillo A, & Veissière S (2017). Culture and depression in global mental health: An ecosocial approach to the phenomenology of psychiatric disorders. Social Science & Medicine, 183, 163–168. 10.1016/j.socscimed.2017.04.034 [PubMed: 28478959]

Kleinheksel AJ, Rockich-Winston N, Tawfik H, & Wyatt TR (2020). Demystifying content analysis. American Journal of Pharmaceutical Education, 84(1), Article 7113. 10.5688/ajpe7113 [PubMed: 32292185]

Kroenke K, Spitzer RL, & Williams JB (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606–613. 10.1046/j.1525-1497.2001.016009606.x [PubMed: 11556941]

Labouliere CD, Vasan P, Kramer A, Brown G, Green K, Rahman M, Kammer J, Finnerty M, & Stanley B (2018). "Zero Suicide"–A model for reducing suicide in United States behavioral healthcare. Suicidologi, 23(1), 22–30. 10.5617/suicidologi.6198 [PubMed: 29970972]

Liu JJ, Reed M, & Girard TA (2017). Advancing resilience: An integrative, multi-system model of resilience. Personality and Individual Differences, 111, 111–118. 10.1016/j.paid.2017.02.007

Mazzuca A, Nagarchi D, Ramanaik S, Raghavendra T, Javalkar P, Rotti S, Bhattacharjee P, Isac S, Cohen A, & Beattie T (2019). Developing a mental health measurement strategy to capture psychological problems among lower caste adolescent girls in rural, South India. Transcultural Psychiatry, 56(1), 24–47. 10.1177/1363461518789540 [PubMed: 30113275]

Mohatt NV, Fok CCT, Burket R, Henry D, & Allen J (2011). Assessment of awareness of connectedness as a culturally-based protective factor for Alaska Native youth. Cultural Diversity & Ethnic Minority Psychology, 17(4), 444–455. 10.1037/a0025456 [PubMed: 21988583]

Moran JR, Fleming CM, Somervell P, & Manson SM (1999). Measuring bicultural ethnic identity among American Indian adolescents: A factor analytic study. Journal of Adolescent Research, 14(4), 405–426. 10.1177/0743558499144002

Mullany B, Barlow A, Neault N, Billy T, Jones T, Tortice I, Lorenzo S, Powers J, Lake K, Reid R, & Walkup J (2012). The family spirit trial for American Indian teen mothers and their children: CBPR rationale, design, methods and baseline characteristics. Prevention Science, 13(5), 504–518. 10.1007/s11121-012-0277-2 [PubMed: 22932743]

Oetting ER, & Beauvais F (1990–1991). Orthogonal cultural identification theory: The cultural identification of minority adolescents. The International Journal of the Addictions, 25(5A-6A), 655–685. 10.3109/10826089109077265

O'Keefe VM, Haroz EE, Goklish N, Ivanich J, Cwik MF, Barlow A, & the Celebrating Life Team. (2019). Employing a sequential multiple assignment randomized trial (SMART) to evaluate the impact of brief risk and protective factor prevention interventions for American Indian youth suicide. BMC Public Health, 19(1), Article 1675. 10.1186/s12889-019-7996-2 [PubMed: 31830933]

Prince-Embury S (2008). The resiliency scales for children and adolescents, psychological symptoms, and clinical status in adolescents. Canadian Journal of School Psychology, 23(1), 41–56. 10.1177/0829573508316592

Reynolds WM (1988). Suicidal ideation questionnaire (SIQ) Professional manual. Psychological Assessment Resources.

Reynolds WM, & Mazza JJ (1999). Assessment of suicidal ideation in inner-city children and young adolescents: Reliability and validity of the suicidal ideation questionnaire-JR. School Psychology Review, 28(1), 17–30. 10.1080/02796015.1999.12085945

Rosenberg M (2015). Society and the adolescent self-image. Princeton University Press.

Snyder CR, Hoza B, Pelham WE, Rapoff M, Ware L, Danovsky M, Highberger L, & Stahl KJ (1997). The development and validation of the Children's Hope Scale. Journal of Pediatric Psychology, 22(3), 399–421. 10.1093/jpepsy/22.3.399 [PubMed: 9212556]

Tavakol M, & Dennick R (2011, June 27). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55. 10.5116/ijme.4dfb.8dfd [PubMed: 28029643]

Ungar M, & Liebenberg L (2011). Assessing resilience across cultures using mixed methods: Construction of the child and youth resilience measure. Journal of Mixed Methods Research, 5(2), 126–149. 10.1177/1558689811400607

Walls ML, Whitbeck LB, Hoyt DR, & Johnson KD (2007). Early-onset alcohol use among Native American youth: Examining female caretaker influence. Journal of Marriage and Family, 69(2), 451–464. 10.1111/j.1741-3737.2007.00376.x

Walls ML, Whitesell NR, Barlow A, & Sarche M (2019). Research with American Indian and Alaska Native populations: Measurement matters. Journal of Ethnicity in Substance Abuse, 18(1), 129–149. 10.1080/15332640.2017.1310640 [PubMed: 28441113]

Weaver LJ, & Kaiser BN (2015). Developing and testing locally derived mental health scales: Examples from North India and Haiti. Field Methods, 27(2), 115–130. 10.1177/1525822X14547191

Whitbeck LB, Adams GW, Hoyt DR, & Chen X (2004). Conceptualizing and measuring historical trauma among American Indian people. American Journal of Community Psychology, 33(3–4), 119–130. 10.1023/B:AJCP.0000027000.77357.31 [PubMed: 15212173]

Whitbeck LB, Yu M, McChargue DE, & Crawford DM (2009). Depressive symptoms, gender, and growth in cigarette smoking among indigenous adolescents. Addictive Behaviors, 34(5), 421–426. 10.1016/j.addbeh.2008.12.009 [PubMed: 19167169]

Whiteside SP, & Lynam DR (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Personality and Individual Differences, 30(4), 669–689. 10.1016/S0191-8869(00)00064-7

Wilson DB, & Lipsey MW (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. Psychological Methods, 6(4), 413–429. 10.1037/1082-989X.6.4.413 [PubMed: 11778681]

Public Significance Statement

Studies addressing the psychometric evaluation of mental health assessment instruments for use in American Indian populations are relatively rare. This study demonstrates how brief qualitative methods can aid in the adaptation and development of psychometrically valid assessment instruments that reflect local understandings of mental health and well-being.

Table 1

List of Priority Changes Youth Experience Because of Having Had Either New Hope or the Elders Resilience Interventions

New hope	Elders resilience intervention
Learning not to get mad	More respect toward peers [a]
Not making bad choices	Wanting to be helpful to other people [a]
How to acknowledge that it is ok to let out your emotions [a]	Sense of self (who you are and where you come from, sense of belonging to family, group of friends, being Apache)
Identifying kinds of emotions [a]	Ask more questions
Their behavior toward their problems and the people in their lives	Eyes light up/Eyes are brighter
Sense of support	Drawing in elder as their own grandma (i.e., treating all elders with respect) [a]
Sense of hope	Interested
Learn how to change negative thoughts to positive thoughts	Engaged
Increase self-esteem	Afraid and shy at the beginning—build confidence overtime
Sense of relief for sharing their story	Confidence (Say some of the words [in Apache], no one will laugh at them)
Feeling connected/belonging	Take better care of themselves, way they dress changes, health behaviors, self-respect [a]
More knowledge	Do something on your own—take action and feel like you can do other activities
Coping skills	They seem happier
Understanding its ok to cry [a]	Hardness becomes more compassionate
Building relationships [a]	Facial expressions become more tender
Better attitudes	Elders help make kids feel like someone cares for them
Outcome of life	Decrease loneliness (creator is always there)
Finding support	Sense of belonging
Realize they have more support than they thought	Better empathy [a]
Using personal skills for positive outlet	
Realize that spiritual is most helpful in prayer	

[a] Changes mentioned that were not covered by any item on the initial assessment battery

Table 2

Changes to the Assessment Battery and Rationale

Original measure	Change & rationale	New measure
Demographics	Retained, slight edits	N/A
Suicide Ideation Questionnaire	Removed three items for lack of local relevance	N/A
Resilience scales	Removed four items, added six locally relevant items, reworded other items, and changed response categories	N/A
Centers for epidemiologic studies of depression	Retained, no changes	N/A
Apache hopefulness scale	Removed and replaced due to items not being relevant, and length	Children's hope scale
Youth risk behavior survey substance use items	Removed and replaced due to existing measure only capturing amount and frequency of use, but not consequences	The WHO alcohol, smoking and substance involvement screening test (ASSIST)
UPPS impulsive behavior scale	Retained, no changes	N/A
Hemingway measure of adolescent connectedness	Removed and replaced. Items did not seem to be capturing the type of community connectedness described during the NGT	Multicultural mastery scale
Voices of Indian teens cultural issues and interest	Retained, no changes	N/A
Rosenberg self esteem scale	Retained, no changes	N/A

Table 3

Index of Local Indicators of Well-Being

Item	Response options			
	Not at all	A little bit	A moderate amount	A lot
I am able to identify what kinds of emotions I am having	0	1	2	3
I feel comfortable crying	0	1	2	3
I want to be helpful to other people	0	1	2	3
I take care of myself	0	1	2	3
I care about the way I dress	0	1	2	3
I take care of my health	0	1	2	3
I care about others	0	1	2	3
I like to understand how others are feeling	0	1	2	3
I have respect for tribal elders	0	1	2	3
I care about the well-being of tribal elders	0	1	2	3
I like helping tribal elders	0	1	2	3

Table 4

Sample Demographics

Characteristics	N	%
Gender (91)		
Female	64	70
Male	27	30
Age (93)		
10–14	47	51
15–19	23	25
20–24	27	18
Behavior (93)		
Ideation	58	62
Attempt	28	30
Binge	7	8

Table 5

Distributions and Cronbach's α of Each Measure

Measures	N	Mean	St. Dev.	Min (possible)	Max (possible)	α
SIQ	93	28.44	20.33	0 (0)	70 (72)	0.96
Mastery	93	33.59	11.65	0 (0)	60 (60)	0.91
Relatedness	93	29.53	10.61	0 (0)	51 (51)	0.88
Emotional reactivity	93	37.98	9.38	0 (0)	60 (60)	0.91
Well-being	93	23.74	5.67	0 (0)	33 (33)	0.82

Table 6

Correlation Matrix

Measures	1	2	3	4	5	6
1. SIQ	—					
2. Mastery	-0.42 ***	—				
3. Relatedness	-0.31 **	0.65 ***	—			
4. Emotional Reactivity	-0.38 ***	0.30 **	0.20	—		
5. Depression	0.68 ***	-0.37 ***	-0.39 ***	-0.46 ***	—	
6. Hope	-0.35 ***	0.57 ***	0.41 ***	0.29 **	-0.44 ***	—
7. Well-being	-0.17	0.52 ***	0.46 ***	0.32 **	-0.24 *	0.51 **

*
$p < .05$.

**
$p < .01$.

***
$p < .001$.